PRINCIPAL COMPONENTS IN RICE RESEARCH

Burton T. Oñate

In rice research, p variates or measurements are usually observed on each of n plants for each of the t treatments. Presently, variability studies are concerned with univariate or one-dimensional space. Oñate and Moomaw (1965)[a] reported the variability of 22 characteristics in rice research. In a recent paper, Oñate (1965)[b] extended these variability studies to 62 characteristics. The behavior of the rice plant to a treatment is generally associated with one or at most two characteristics or factors taken simultaneously. The reaction also can be studied with three or more characteristics observed on the rice plant. Principal component analysis will help economize in the number of characteristics which can be substantial. In addition, the p original variates (X_{ij}; i = 2, ..., p; j = 1, 2, ..., n) are transformed into p new component (Z_i; i = 1, 2, ..., p) such that the Z_i's are independent and the first two or three of the Z_i's will account for 85 to 95 percent of total variation. The correlation matrix (r) is obtained and the latent roots and vectors are solved from the characteristic equation.

$$\left| r - \lambda I \right| = O$$

where

r is the correlation matrix,
λ is the latent root and there are p values of λ,

and

I is the identity matrix.

For each λ_i, there is a solution vector (w_i) which will serve as coefficients of the standardized X's. The first principal component is given by

[a] OÑATE, B.T. and J.C. MOOMAW, Variability of Characteristics in Rice Research. The Phil. Agric., XLIX (1), June 1966, 8-15.
[b] OÑATE, B.T. Statistical Problems in Rice Research. The International Rice Research Institute, October 1965.

$$Z_l = \sum_{k=1}^{p} w_{lk} x_k$$

where the w_{lk} relates to λ_l and the variance of Z_l is λ_l. Similar relationships are obtained for the other Z's. Other properties of latent roots are

$$\sum_{i=1}^{p} \lambda_i = p$$

and

$$\prod_{i=1}^{p} \lambda_i = \left| r \right| .$$

The analysis for this study was programmed in the TOSBAC 3400 electronic computer.

Description of Experiments

The data were collected by the Department of Plant Physiology, IRRI during the 1964 dry season. Principal component analysis was applied to three sets of data, namely:

I. variety x. nitrogen. Three levels of nitrogen 0, 50 and 150 kg/ha and spacing of 25 cm. x 25 cm. were studied.
II. variety x spacing. With 50 kg/ha the plants were spaced at 10 cm. x 10 cm., 25 cm. x 25 cm. and 50 cm. x 50 cm.
III. variety x nitrogen x spacing. The data for 1) and 2) were used for a combined analyses.

The 10 varieties used in I to III in increasing growth duration are as follows:

(1) Fujisaka (F)
(2) Hakkoda
(3) Sukhwel
(4) Nang Quot
(5) Cuon Trau
(6) Tjeremas

(7) Peta
(8) 59-368
(9) Acheh
(10) Serayap

The value of p = 8 plant characteristics which are given the following descriptions:

X_1 = yield per plant in grams,
X_2 = number of mature panicles per plant,
X_3 = spikelet number per panicle,
X_4 = number of grains per panicle,
X_5 = sterility per plant in percent,
X_6 = weight of straw per plant in grams,
X_7 = tiller number per plant,
X_8 = plant height in cm.

Principal Components Analysis—(variety x nitrogen)

The (8 x 8) correlation matrix r is given in table 1a. The eigenvalues or latent roots and their probabilities for the eight principal components are given in table 1b. The first principal component Z_1 has an eigenvalue of 4.2 which represents 53 percent of the total variability. The second principal component Z_2 has a value of 2.44 with a probability of 30.5 while the third component Z_3 has a variance of 0.87 which represents 10.9 percent of the overall variability. The first two principal components Z_1, Z_2 accounted for a total of 83 percent and the first three components accounted for a total of 94 percent of the total variability. In this case, we can say that the eight variables $x_1, x_2, \ldots, x_8$ can be represented by the first three principal components Z_1, Z_2 and Z_3.

The three principal components as expressed in terms of the standardized variates are as follows:

$$Z_1 = 0.42x_1 - 0.10x_2 0.45x_3 + 0.45x_4 - 0.11x_5 + 0.43x_6 - 0.15x_7 + 0.43x_8$$

$$Z_2 = 0.24x_1 x 0.60x_2 - 0.02x_3 - 0.16x_4 0.37x_5 + 0.23x_6 + 0.58x_7 + 0.17x_8$$

$$Z_3 = -.30x_1 - 0.23x_2 + 0.29x_3 - 0.06x_4 + 0.06x_5 - 0.02x_6 - 0.22x_7 + 0.16x_8$$

The first principal component Z_1 may be termed as indicator of 'size' or 'volume'. Upon inspection of the weights (w_{1k}) the largest values are attached to yield (x_1), spikelet number per panicle (x_3), number of grains per panicle (x_4), weight of straw per plant in grams (x_6), and plant height (x_8). The weights are almost equal ranging from 0.42 for yield (x_1) to 0.45 for spikelet number per panicle (x_3). Since the correlation between x_3 and x_4 is high, only one of these variables may be retained in subsequent analysis.

The second component Z_2 accounts for 30.5 percent of the variance. The largest weights are given by the number of mature panicles (x_2) and the number of tillers (x_7). These two variables may be associated with 'shape' or 'density'. The correlation between x_2 and x_7 is $+0.96$. This implies that in subsequent analysis we can use either the number of mature panicles or the number of tillers.

The third principal component Z_3 which accounts for almost 11 percent of the total variability is clearly associated with sterility (x_5); the correlation coefficient is equal to $+0.78$.

In summary, Z_1 is highly correlated with five (or four) of the original variables, Z_2 to two (or one) and Z_3 to one variable (table 1c).

TABLE 1a

CORRELATION MATRIX OF THE EIGHT CHARACTERISTICS.
(Variety x Nitrogen)

Characteristics	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
x_1	1.00	0.22	0.69	0.71	-0.18	-0.87	0.11	0.80
x_2		1.00	-0.26	-0.39	0.43	0.12	0.96	0.01
x_3			1.00	0.91	-0.01	0.75	-0.34	0.78
x_4				1.00	-0.39	0.68	-0.47	0.69
x_5					1.00	-0.02	0.44	0.06
x_6						1.00	0.04	0.90
x_7							1.00	-0.08
x_8								1.00

TABLE 1b

EIGENVALUES AND PERCENT VARIATION FOR THE EIGHT PRINCIPAL COMPONENTS.

(Variety x Nitrogen)

Principal component	Eigen value	Percent variation	Accumulated percentage
Z_1	4.20	52.6	52.6
Z_2	2.44	30.5	83.1
Z_3	0.87	10.9	94.0
Z_4	0.27	3.4	97.4
Z_5	0.09	1.1	98.5
Z_6	0.08	1.0	99.5
Z_7	0.03	0.3	99.8
Z_8	0.01	0.2	100.0

TABLE 1c

CORRELATION BETWEEN THE PRINCIPAL COMPONENTS AND THE EIGHT CHARACTERISTICS.

(Variety x Nitrogen)

Principal component	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
Z_1	0.85	-0.21	0.92	0.93	-023	0.88	-0.31	0.89
Z_2	0.37	0.94	-0.03	-0.25	0.58	0.36	0.91	0.27
Z_3	-0.28	-0.21	0.27	-0.06	0.78	-0.02	-0.20	0.15
Z_4	-0.01	-0.10	-0.26	-0.25	-0.03	0.21	-0.11	0.27
Z_5	-0.21	-0.02	0.04	0.05	-0.06	0.15	0.11	-0.02

Upon examination of table 1c, one will note that the correlation of Z_4 with the original x's are all low. Similarly, this statement holds true with Z_5. These evidences show that the first three components would suffice to explain almost all of the variations inherent in the x's.

The results for the variety x nitrogen experiments are given in detail in tables 1a, 1b and 1c. A summary of the results for the three sets of experiments is given in table 2. Yield number of spikelets, number of grains and straw are common to Z_1;

number of panicles and number of tillers to Z_2; while sterility is identified with Z_3 for the three experiments I, II and III. Height appeared in Z_1 for experiment I and with Z_2 in experiments II and III. In all experiments Z_1, Z_2 and Z_3 accounted for 93 or 94 percent of total variation.

Since number of filled grains is a component of number of spikelet (filled + unfilled grains), the correlation between these two characteristics ranged from +0.86 to +0.91. Similarly, the correlation between numbers of panicles and tillers is +0.97. In subsequent analysis, only one of these characteristics may be retained, filled grain or spikelet in Z_1 and panicle or tiller in Z_2.

Sterility is derived as a ratio of unfilled grain to number of spikelet. The observed correlation between sterility and grain is -.28 to -.37 while between sterility and spikelet the range is -.01 to -.05. These correlations are very low to be of any practical significance. Thus, sterility appears to be a strong defining contrast in rice research.

Interaction (Variety x Nitrogen)

Since there were a total of six plants for each variety by treatment interaction, it may be worthwhile to see the reaction of the plants in each (variety x nitrogen) interaction for each of the (Z_1, Z_2), (Z_1, Z_3) axes. There will be a total of thirty (10 varieties x 3 nitrogen levels) points. It was noted that the principal components by plants and by replications for each treatment combination are stable. Thus the mean value of the Z's was computed and graphed in Fig. 1. In general, the shift of increasing nitrogen for each variety is verticle or along the Z_2 axis which implies that as nitrogen is increased the effect is primarily to increase the number of mature panicles and tillers. Note that variety Acheh (A) shifted along the Z_2 axis from ON to 5ON and changed direction along the positive Z_1 axis from 5ON to 15ON, this reaction indicates a shift along two axes with different levels.

While the major shift of increasing nitrogen is along the Z_2 axis, some of the varieties also indicated a shift in both Z_1 and Z_2 axes. As the growth duration increased there is an increase of values in the Z_1 axis. These results imply that varieties with long duration is generally associated with higher yield, more spikelet number per panicle or more grains per panicle, higher weight of straws and taller plants. An entirely different picture may be obtained if instead of grams/plant, we use grams/day/plant in the analysis.

The reactions on the (Z_1, Z_3) and (Z_2, Z_3) axes are available.